Robust Parameter Estimation For Mixture Model

Saldju Tadjudin
Netcom Systems, Inc.
20550 Nordhoff Street
Chatsworth, CA 91311
Phone (818) 885-2179
Fax (818) 709-0117
Saldju_Tadjudin@netcomsystems.com

David A. Landgrebe
School of Electrical and Computer Engineering
Purdue University
West Lafayette, IN 47907-1285
Phone (765) 494-3486
Fax (765) 494-3358
Landgreb@ecn.purdue.edu

 This work originally appeared In the *IEEE Transactions on Geoscience and Remote Sensing, Vol. 38, No. 1, pp. 439-445, January 2000*.

ABSTRACT

In pattern recognition, when the ratio of the number of training samples to the dimensionality is small, parameter estimates become highly variable, causing the deterioration of classification performance. This problem has become more prevalent in remote sensing with the emergence of a new generation of sensors with as many as several hundred spectral bands. While the new sensor technology provides higher spectral and spatial resolution, enabling a greater number of spectrally separable classes to be identified, the needed labeled samples for designing the classifier remain difficult and expensive to acquire. Better parameter estimates can be obtained by exploiting a large number of unlabeled samples in addition to training samples using the expectation maximization algorithm under the mixture model. However, the estimation method is sensitive to the presence of statistical outliers. In remote sensing data, miscellaneous classes with few samples are often difficult to identify and may constitute statistical outliers. Therefore, we propose to use a robust parameter estimation method for the mixture model. The proposed method assigns full weight to training samples, but automatically gives reduced weight to unlabeled samples. Experimental results show that the robust method prevents performance deterioration due to statistical outliers in the data as compared to the estimates obtained from the direct EM approach.

Work leading to this paper was supported in part by NASA under Grant NAG5-3975 and the Army Research Office under Grant DAAH04-96-1-0444.

INTRODUCTION

In a mixture model, data are assumed to consist of two or more distributions mixed in varying proportions. For remote sensing applications, it is a common practice to consider several "spectral subclasses" within each "information class" or ground cover type. Each of such spectral subclasses is considered to be multivariately normally distributed and classification is then performed with respect to the spectral subclasses. Under this model, we can regard remote sensing data as a mixture model fitted with normally distributed components. To estimate the model parameters in a mixture, a common approach is to apply the expectation maximization (EM) algorithm, which is an iterative method for numerically approximating the maximum likelihood (ML) estimates of the parameters in a mixture model. Alternatively, it can be viewed as an estimation problem involving incomplete data in which each unlabeled observation is regarded as missing a label of its origin [1].

In [2], the EM algorithm has been studied and applied to remote sensing data. It was shown that by assuming a mixture model and using both training samples and unlabeled samples in obtaining the class distribution estimates, the classification performance can be improved. Also, the Hughes phenomenon [3] can be delayed to a higher dimensionality and hence more features can be used to obtain better performance. In addition, the parameter estimates represent the true class distributions more accurately. However, the unrepresented pixel classes have been dealt with by rejection using a chi-square threshold. This method can be viewed as a hard decision. Unfortunately, a suitable threshold value is difficult to select and is usually arbitrary. Consequently, "useful" pixels might be rejected as outliers. We propose to use a robust method to estimate the mean vector and covariance matrix for classifying multispectral data under the mixture model. This approach assigns full weight to the training samples, but automatically gives reduced weight to unlabeled samples. Therefore, it avoids the risk of rejecting useful pixels while still limiting the influence of outliers in obtaining the ML estimates of the parameters. In the next section, the EM algorithm is reviewed and discussed.

EXPECTATION MAXIMIZATION ALGORITHM

The Expectation Maximization (EM) algorithm is an iterative method for numerically approximating the maximum likelihood (ML) estimates of the parameters in a mixture model. Alternatively, it can be viewed as an estimation problem involving

incomplete data in which each unlabeled observation in the mixture is regarded as missing its label.

Under the mixture model, the distribution of the data $x \in \Re^p$ is given as:

$$f(x;\Phi) = \sum_{i=1}^{L} \alpha_i f_i(x;\phi_i)$$

where $\alpha_1,\ldots,\alpha_L$ are the mixing proportions, f_i is the component density parameterized by ϕ_i and L is the total number of components. The mixture density f is then parameterized by $\Phi = (\alpha_1,\ldots,\alpha_L,\phi_1,\ldots,\phi_L)$.

Under the incomplete data formulation, each unlabeled sample x is considered as the labeled sample y with its class origin missing. Therefore, we can denote $y = (x,i)$ where $i = 1\cdots L$ indicates the sample origin. Let $g(x|\Phi)$ be the probability density function (pdf) of the incomplete data $x = (x_1,\ldots,x_n)$ and $f(y|\Phi)$ be the pdf of the completely labeled data $y = (y_1,\ldots,y_n)$. The maximum likelihood estimation then involves the maximization of the log likelihood of the incomplete data $L(\Phi) = logg(x|\Phi)$. The estimation is complicated by the fact that the sample origin is missing. Hence, the EM algorithm uses the relationship between $f(y|\Phi)$ and $g(x|\Phi)$ to maximize the incomplete data log-likelihood $L(\Phi) = logg(x|\Phi)$. Using an iterative approach, the EM algorithm obtains the maximum likelihood estimates by starting with an initial estimate Φ^0 and repeating the following two steps at each iteration:

E-Step) Determine $Q\left(\Phi\middle|\Phi^c\right) = E\left\{log\ f(y|\Phi)\middle|x,\Phi^c\right\}$

M-Step) Choose $\Phi^+ = argmaxQ\left(\Phi\middle|\Phi^c\right)$

The next and current values of the parameters are denoted by the superscripts "+" and "c" respectively. The algorithm begins with an initial estimate. It has been shown that under some relatively general conditions the iteration converges to ML estimates, at least locally. Since the convergence is only guaranteed to a local maximum, the algorithm usually must be repeated from various initial points. However, the training samples, if available, can provide good initial estimates.

Assume that $y = (y_1,\ldots,y_{m_i})$ are the m_i training samples from class i. Also, there are L Gaussian classes and a total of n unlabeled samples denoted by $x = (x_1,\ldots,x_n)$. The

parameter set Φ then contains all the prior probabilities, mean vectors and covariance matrices. The EM algorithm can then be expressed as the following iterative equations [4]:

E-Step:

$$\tau_{ij}^{c} = \tau_i\left(x_j|\phi_i^{c}\right) = \alpha_i^{c} f_i\left(x_j|\phi_i^{c}\right) \Big/ \sum_{t=1}^{L} \alpha_t^{c} f_t\left(x_j|\phi_t^{c}\right) \tag{1}$$

where τ_{ij}^{c} is the posterior probability that x_j belongs to class i.

M-Step:

$$\alpha_i^{+} = \sum_{j=1}^{n} \tau_{ij}^{c} / n \tag{2}$$

$$\mu_i^{+} = \frac{\sum_{j=1}^{m_i} y_{ij} + \sum_{j=1}^{n} \tau_{ij}^{c} x_j}{m_i + \sum_{j=1}^{n} \tau_{ij}^{c}} \tag{3}$$

$$\Sigma_i^{+} = \frac{\sum_{j=1}^{m_i} \left(y_{ij} - \mu_i^{+}\right)\left(y_{ij} - \mu_i^{+}\right)^T + \sum_{j=1}^{n} \tau_{ij}^{c}\left(x_j - \mu_i^{+}\right)\left(x_j - \mu_i^{+}\right)^T}{m_i + \sum_{j=1}^{n} \tau_{ij}^{c}} \tag{4}$$

There are several factors affecting the convergence of the EM algorithm to the maximum likelihood estimates. First of all, the selection of training samples as initial estimates can affect the convergence to a great extent. In this work, the training set is assumed to provide a good initial estimate. Another factor that affects the performance of the EM algorithm is the presence of statistical outliers. Assume that the number of components have been decided and given by the training set. Statistical outliers are defined as those observations that are substantially different from the distributions of the mixture components. As indicated by Eq. (1) through Eq. (4), the EM algorithm assigns each observation to one of the components with the sample's posterior probability as its weight. Even though an outlying sample is inconsistent with the distributions of all the defined components, it may still have a large posterior probability for one or more of the components. As a result, the iteration converges to erroneous solutions.

The problem of outliers is not uncommon in practical applications. In remote sensing, a scene usually contains pixels of unknown origin which form "information

noise". For example, in an agricultural area, there could be pixels belonging to houses, trees or rural roads. The statistical distributions of these pixels may be significantly different from those of training classes and constitute statistical outliers. Unfortunately, these outlying pixels are usually scattered throughout the scene and are small in number. Consequently, identifying these pixels could be a tedious task. A common approach to eliminate those pixels in the EM algorithm is to apply a chi-square threshold test. In other words, pixels whose distances are greater than the threshold value are considered as outliers and are subsequently excluded from updating the estimates. The chi-square threshold T_α for a given probability α is defined as the squared distance between the sample $x \in \Re^p$ and the mean vector for class i based on the chi-square distribution as shown in the following [11]:

$$\Pr\left\{ x \middle| (x - m_i)^T \Sigma_i^{-1} (x - m_i) \le T_\alpha \right\} = \alpha .$$

The thresholding approach can be regarded as performing a hard decision to eliminate outlying samples before initiating the EM algorithm. A suitable threshold value is often difficult to select and is usually arbitrary. Consequently, "useful" pixels might be rejected as statistical outliers. In particular, as dimensionality increases, most pixels might be considered as outliers. An alternative would be to assign a different weight to each pixel and use all available unlabeled pixels for updating the statistics. This method can be regarded as applying a soft decision. In the next section, the robust EM equations will be discussed and modified to process remote sensing data.

ROBUST ESTIMATION

The robust estimation of model parameters was first developed as Huber [5] proposed a theory of robust estimation of a location parameter using M-estimates in a non-mixture context. It was later extended to the multivariate case by taking an elliptically symmetric density and then associating it with a contaminated normal density [6]. Campbell [7] derived the M-estimates for the mixture density and obtained an EM-like algorithm but with a weight function assigned to each pixel as a measure of typicality. The outlier problem in remote sensing has been addressed in [8]. The author proposed a modified M-estimation of the parameters to deal with the situation when the training samples of a certain information class contain samples of other classes. This is typical for a mixture model. The modified M-estimates were shown to be robust with respect to the contamination in the training samples as compared to the least-square estimates. However, the use of unlabeled samples in updating statistics was not addressed. This section will describe the method of

robust EM algorithm following the discussion in [7], and adapting the approach for remote sensing data.

The EM algorithm first estimates the posterior probabilities of each sample belonging to each of the component distributions, and then computes the parameter estimates using these posterior probabilities as weights. With this approach, each sample is assumed to come from one of the component distributions, even though it may greatly differ from all components. The robust estimation attempts to circumvent this problem by including the typicality of a sample with respect to the component densities in updating the estimates in the EM algorithm.

To incorporate a measure of typicality in the parameter estimation of the mixture density, each component density $f_i(x|\mu_i, \Sigma_i)$ for $x \in \Re^p$ is assumed to be a member of the family of p-dimensional elliptically symmetric densities with mean vector μ_i and covariance matrix Σ_i [7]:

$$|\Sigma_i|^{-1/2} f_S\{\delta_i(x;\mu_i,\Sigma_i)\}$$

where $\delta_i^2 = (x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)$. Typically, $f_S(\delta_i)$ is assumed to be the exponential of some symmetric function $\rho(\delta_i)$:

$$f_S(\delta_i) = exp\{-\rho(\delta_i)\}.$$

Then, the likelihood parameter estimation for these component densities can be obtained by applying the expectation and maximization steps. Denoting the current and future parameter values by the superscripts "c" and "+", the iterative equations are derived as [7]:

$$\alpha_i^+ = \sum_{j=1}^{n} \tau_{ij}^c / n$$

$$\mu_i^+ = \sum_{j=1}^{n} \tau_{ij}^c w_{ij}^c x_j \Big/ \sum_{j=1}^{n} \tau_{ij}^c w_{ij}^c$$

$$\Sigma_i^+ = \sum_{j=1}^{n} \tau_{ij}^c w_{ij}^+ (x_j - \mu_i^+)(x_j - \mu_i^+)^T \Big/ \sum_{j=1}^{n} \tau_{ij}^c$$

where $w_{ij} = \psi(\delta_{ij})/\delta_{ij}$ is the weight function and $\psi(\delta_{ij}) = \rho'(\delta_{ij})$ is the first derivative of $\rho(\delta_{ij})$. To limit the influence of large atypical samples, the covariance estimator is modified to be:

$$\Sigma_i^+ = \sum_{j=1}^{n} \tau_{ij}^c w_{ij}^{+2} (x_j - \mu_i^+)(x_j - \mu_i^+)^T \Big/ \sum_{j=1}^{n} \tau_{ij}^c w_{ij}^{+2}.$$

The weight function has been chosen to be $\psi(s)/s$ where $s = \delta_{ij}$. A popular choice of $\psi(s)$ is the Huber's ψ -function that is defined by $\psi(s) = -\psi(-s)$ where for $s > 0$

$$\psi(s) = \begin{cases} s & 0 \le s \le k_1(p) \\ k_1(p) & s > k_1(p) \end{cases}$$

for an appropriate choice of the "tuning" constant $k_1(p)$, which is a function of the dimensionality p. This selection of $\psi(s)$ gives:

$$\rho(s) = \begin{cases} \frac{1}{2}s^2 & 0 \le s \le k_1(p) \\ k_1(p)s - \frac{1}{2}k_1^2(p) & s > k_1(p) \end{cases}.$$

The value of the tuning constant is a function of dimensionality. It also depends on the amount of contamination in the data that is usually not known. Since the training samples are representative of the classes, it is desirable that they are given more emphasis in the updates of the estimates. Therefore, in the proposed approach, the training samples are assigned unit weight. To do so, the value of $k_1(p)$ is defined to be

$$k_1(p) = max(\hat{d}_{ij})$$

where $\hat{d}_{ij}^2 = (y_{ij} - \mu_i)^T \Sigma_i^{-1} (y_{ij} - \mu_i)$ and y_{ij} is the training sample j from class i. In other words, the tuning constant is selected such that the training samples are given unit weight and the weights for the unlabeled samples are inversely proportional to the square root of their distances to the class mean. Therefore, the weight assigned to each sample can be expressed as:

$$w_{ij} = \begin{cases} 1 & d_{ij} \le max(\hat{d}_{ij}) \\ max(\hat{d}_{ij})/d_{ij} & max(\hat{d}_{ij}) < d_{ij} < \infty \end{cases}$$

where $d_{ij}^2 = (x_j - \mu_i)^T \Sigma_i^{-1} (x_j - \mu_i)$ is the squared distance of unlabeled samples x_j. The iterative equations for the mean and covariance estimates can then be expressed as:

$$\mu_i^+ = \frac{\sum_{j=1}^{m_i} y_{ij} + \sum_{j=1}^{n} \tau_{ij}^c w_{ij}^c x_j}{m_i + \sum_{j=1}^{n} \tau_{ij}^c w_{ij}^c}$$

$$\Sigma_i^+ = \frac{\sum_{j=1}^{m_i} (y_{ij} - \mu_i^+)(y_{ij} - \mu_i^+)^T + \sum_{j=1}^{n} \tau_{ij}^c w_{ij}^{+2} (x_j - \mu_i^+)(x_j - \mu_i^+)^T}{m_i + \sum_{j=1}^{n} \tau_{ij}^c w_{ij}^{+2}}.$$

EXPERIMENTAL RESULTS

In the following experiments, we compare the performance of quadratic classifiers using the parameters estimated from training samples alone (ML), the EM algorithm (EM) and the proposed robust algorithm (REM).

Experiments 1 through 4 are performed using a portion of an AVIRIS data set taken over NW Indiana's Indian Pine test site in June 1992. The scene contains four information classes: corn-no till, soybean-no till, soybean-min till and grass. By visual inspection of the image, the list of these ground cover types is assumed to be exhaustive. A total of 20 channels from the water absorption and noisy bands (104-108, 150-163, 220) are removed from the original 220 spectral channels, leaving 200 spectral features for the experiments. The test data and the ground truth map are shown in Fig. 1. The number of labeled samples in each class is shown in Table 1. Due to the limited labeled samples, we select the number of spectral channels at 10, 20, 50, 67 and 100. These channels are selected by sampling the spectral range at fixed interval. The training samples are randomly selected and the remaining labeled samples are used for testing. The algorithms are repeated for 10 iterations and the classification is performed using the Gaussian maximum likelihood classifier. The maximum likelihood (ML) method using only the training samples to estimate the parameters is denoted as ML in the following experiments.

Class Names	No. of Labeled Samples
Corn-no till	910
Soybean-no till	638
Soybean-min till	1421
Grass	618

Table 1. Class Description for AVIRIS Data.

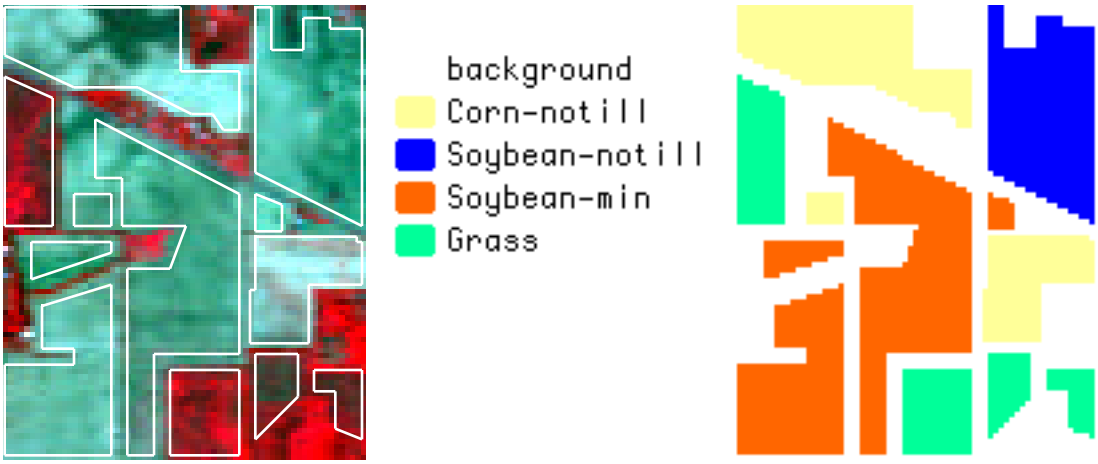

Fig. 1. A portion of AVIRIS Data and Ground Truth Map (Original in Color).

Experiment 1

The first experiment is intended to compare EM and REM without outliers in the data. To obtain data without outliers, we generate synthetic data using the statistics computed from the labeled samples of the four classes. A total of 2000 test samples per class is generated, 500 of which are used as the training samples. Since the training samples are selected at random, the experiment is repeated 5 times and the mean classification accuracy is recorded. The mean accuracy is shown in Fig. 2.

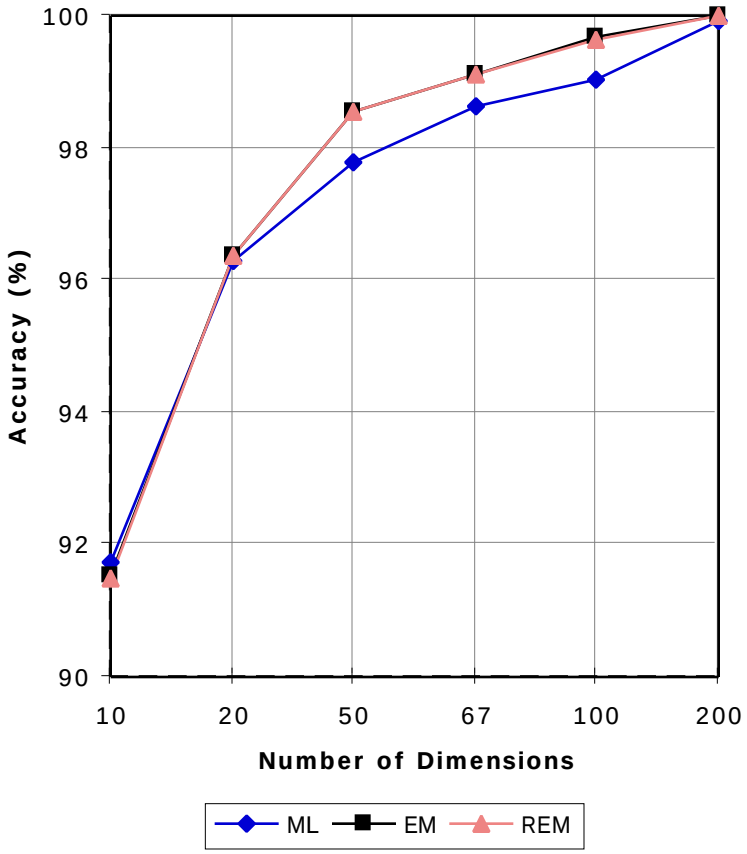

Fig. 2. Mean Accuracy for Experiment 1
with 500 Training Samples and 1500 Test Samples.

The results show that when no outliers are present in the data, the EM and REM algorithms have similar performance and both result in a better performance than the maximum likelihood classifier using the training samples alone. Since there are many design samples available, the best performance is obtained at 200 features.

Experiment 2

In this experiment, the synthetic data from the Experiment 1 is used with the exception that only 250 training samples are selected for each class. The number of test samples is kept at 1500. Again, no outliers are present in the data. The results are shown in Fig. 3.

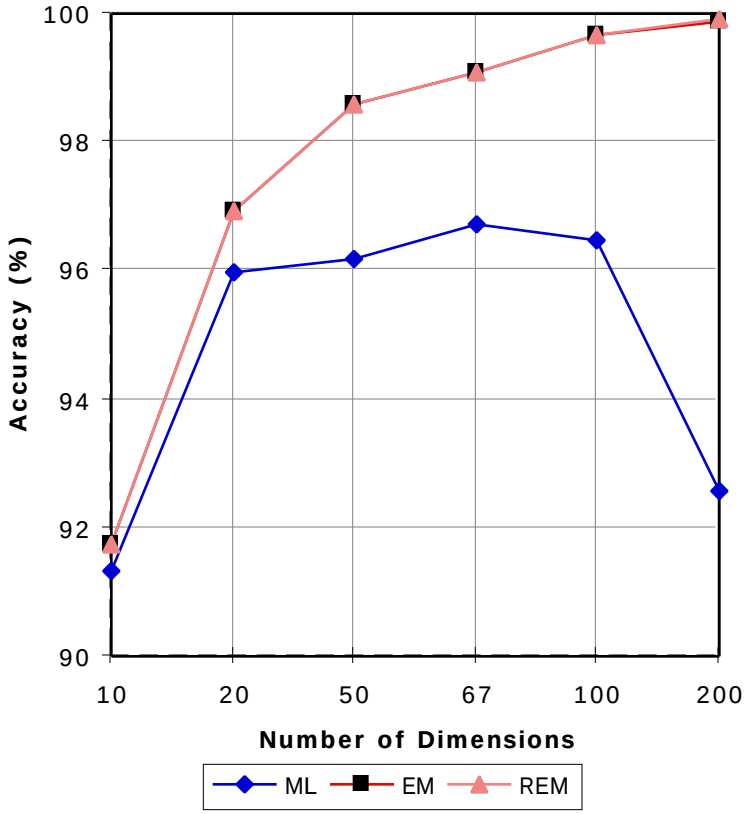

Fig. 3. Mean Accuracy for Experiment 2
with 250 Training Samples and 1500 Test Samples.

Since fewer training samples are used, the performance of the maximum likelihood classifier (ML) using the training samples alone deteriorates. The decline is particularly obvious at higher dimensionality. Compared to the previous experiment, the accuracy has dropped 7% at 200 features. However, when unlabeled samples are used for the mixture model, the performance remains stable even when the number of training samples declines. The results again show that when no outliers are present in the data, the EM and REM algorithms have comparable performance and both achieve better classification accuracy than the ML classifier without using additional unlabeled samples.

Experiment 3

The previous experiment is repeated with only 400 test samples generated for each class. The number of training samples per class is 250. Again, no outliers are present in the data. The results are shown in Fig. 4. Compared to the results from two previous experiments in which many more unlabeled samples were used, the classification results

for all three methods deteriorate in this experiment. This deterioration is manifested as the Hughes phenomenon. Hence, the likelihood parameter estimation for the mixture model is shown to be affected by the number of unlabeled samples relative to dimensionality. Specifically, it implies that 650 samples are still inadequate to characterize these 200-dimensional Gaussian distributions. The results again indicate that without outliers, the EM and REM algorithms have comparable performance and both have better classification accuracy than the ML classifier without using additional unlabeled samples.

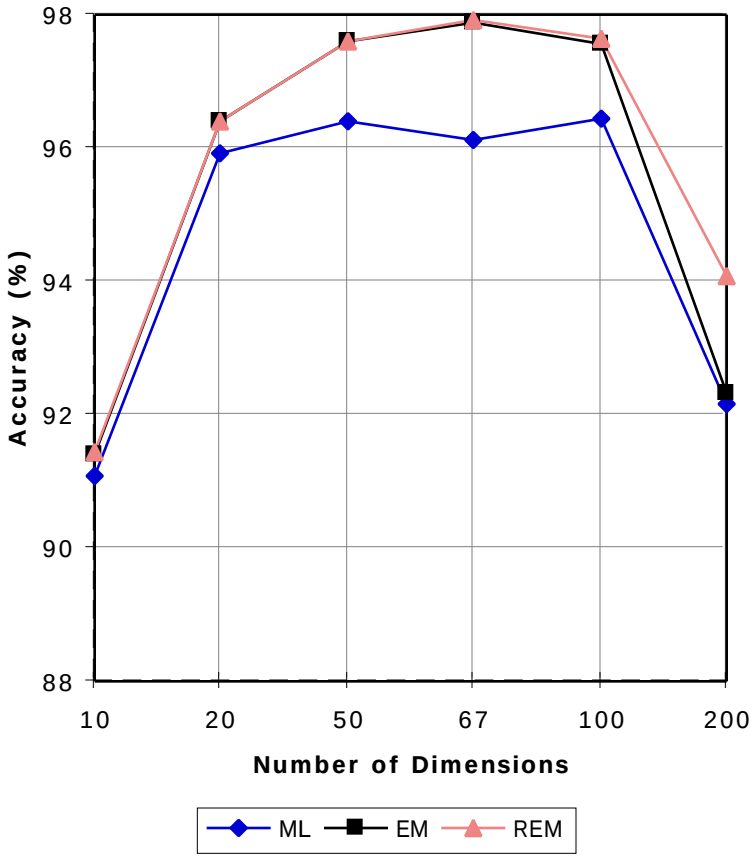

Fig. 4. Mean Accuracy for Experiment 3
with 250 Training Samples and 400 Test Samples.

Experiment 4

This experiment is conducted using the real samples from the data. Again, since all four classes are represented by the training samples, the classes are assumed to be exhaustive. As indicated in Table 1, the number of labeled samples is small. To retain enough test samples, only about 200 training samples are chosen for each class. Due to the limited labeled sample size, to obtain reasonably good initial estimates for comparing the EM and REM algorithms, the number of spectral channels are selected at 10, 20, 50, 67 and 100. These spectral features are again chosen by sampling the spectral channels at fixed intervals. Fig. 5 shows the classification results at the selected dimensions.

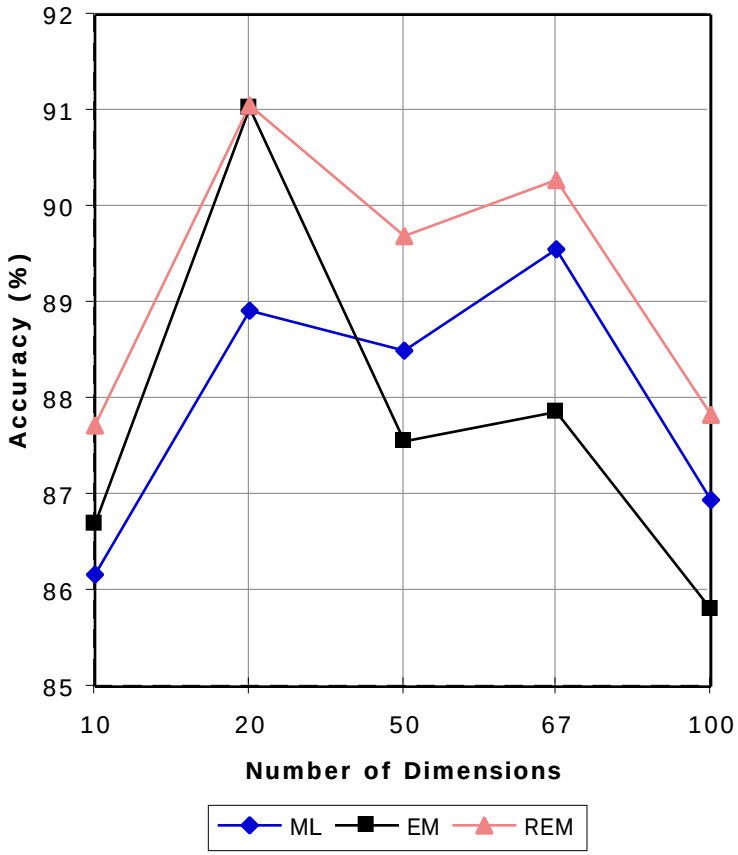

Fig. 5. Accuracy for Experiment 4 using AVIRIS Data.

The results show that the REM algorithm performs better than the ML and EM methods. This demonstrates that although it is assumed that the scene contains no outliers, there are some outlying pixels that were not identified. This further justifies the motivation

of using a robust parameter estimation method for the mixture model. The results also show that all methods exhibit the Hughes phenomenon. As discussed previously, the decline in performance at high dimensionality is caused by the limited number of unlabeled samples available in the image.

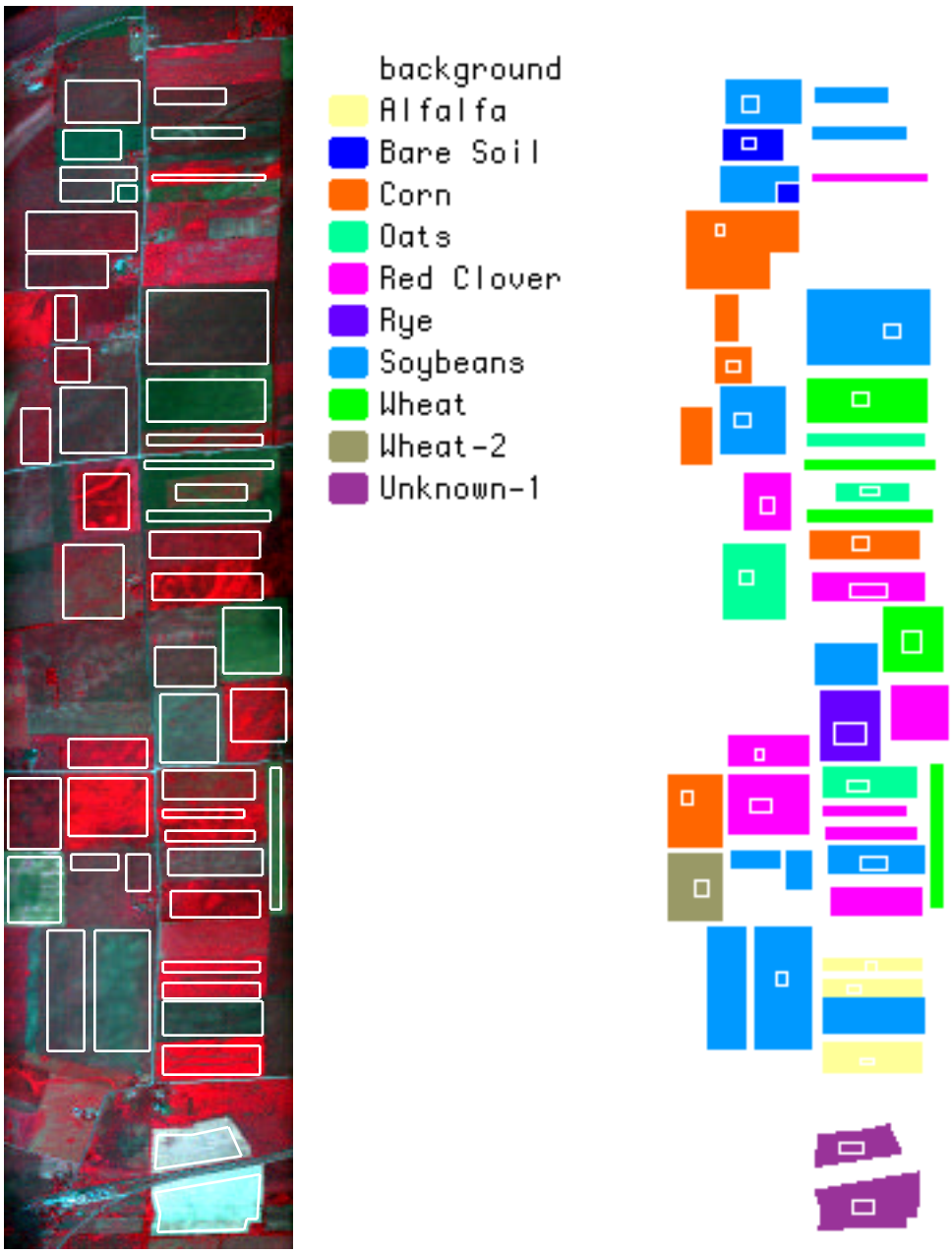

Fig. 6. Flightline C1 Image and Ground Truth Map (Original in Color).

Experiment 5

This experiment is conducted using a data set designated Flightline C1 (FLC1), which is a 12-band multispectral data taken over Tippecanoe County, Indiana by the M7 scanner [10] in June, 1966. The data and the ground truth map are shown in Fig. 6. The training fields are marked in the ground truth map. The number of labeled samples and training samples in each class is shown in Table 2. The parameters are estimated using the training samples alone, the EM algorithm with various threshold settings, and the REM algorithm. For the EM algorithm, two chi-square threshold values (1% and 5%) are applied for comparison. The classification results are plotted in Fig. 7.

Class Names	No. of Labeled Samples	No. of Training Samples
Alfalfa	3375	156
Bare Soil	1230	90
Corn	10625	331
Oats	5781	306
Red Clover	12147	614
Rye	2385	408
Soybeans	25133	631
Wheat	7827	340
Wheat-2	2091	120
Unknown-1	4034	322

Table 2. Class Description for Flightline C1 Data.

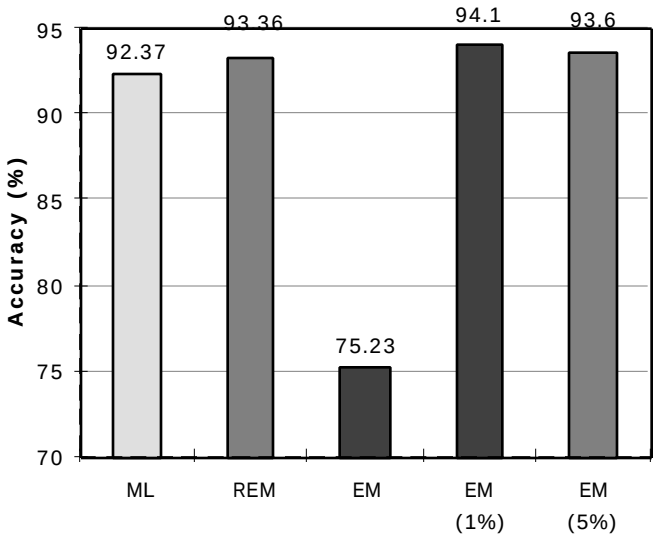

Fig. 7. Classification Results for Flightline C1 Data.

The entire Flightline C1 image contains classes with a few pixels such as rural roads, farmstead and water that are not included in the training set. There may be other unknown classes that are not identified in the ground truth information. Therefore, it is highly likely that statistical outliers are present in the image. This is confirmed by experimental results. The performance of the EM algorithm is significantly lower than those of ML, REM and EM with thresholding. Again, the experiment demonstrates that REM has similar performance as EM with thresholding, but without the need of setting a threshold.

CONCLUSION

In this paper, a robust method for parameter estimation under the mixture model (REM) is proposed and implemented for classifying multispectral data. This work is motivated by the fact that a multispectral image data set usually contains pixels of unknown classes which can be time-consuming to identify. These pixels of unknown origin may have density distributions quite different from the training classes and constitute statistical outliers. Without a list of exhaustive classes for the mixture model, the expectation maximization (EM) algorithm can converge to erroneous solutions due to the presence of statistical outliers. This problem necessitates a robust version of the EM algorithm that includes a measure of typicality for each sample.

The experimental results have shown that the proposed robust method performs better than the parameter estimation methods using the training samples alone (ML) and the EM algorithm in the presence of outliers. When no outliers are present, the EM and REM have similar performances and both are better than the ML approach. Specifically, when there are many unlabeled samples available, the EM and REM algorithms can mitigate the Hughes phenomenon since they utilize unlabeled samples in addition to the training samples. When the number of unlabeled samples are limited, both EM and REM methods exhibit the Hughes phenomenon, but still achieve better classification accuracy than the ML approach at lower dimensionality. Despite the promising results, the proposed REM algorithm has several limitations. Since the weight function in the REM algorithm is based on class statistics, the initial parameter estimates are important in determining the convergence. In particular, a good covariance estimate requires a sufficient number of training samples. When the number of training samples is close to or less than the dimensionality, the covariance estimate becomes poor or singular and the EM or REM algorithm cannot be applied. This necessitates the use of a covariance estimation method for limited training samples. Further details of this algorithm can be found in [9].

REFERENCES

[1] A.P. Dempster, N.M. Laird, D.B. Rubin, "Maximum likelihood estimation from incomplete data via EM algorithm," J. R. Statist. Soc., Vol. B39, pp. 1-38, 1977.

[2] B. Shahshahani and D. A. Landgrebe, Classification of Multi-spectral Data by Joint Supervised-Unsupervised Learning, Purdue University, West Lafayette, IN, TR-EE 94-1, January 1994.

[3] G.F. Hughes, "On the mean accuracy of statistical pattern recognizers," IEEE Trans. Inform. Theory., Vol. IT-14, pp. 55-63, 1968.

[4] R.A. Redner, H.F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," SIAM Review, Vol. 26, No. 2, pp. 195-239, 1984.

[5] P.J. Huber, "Robust estimation of a location parameter," Ann. Math. Statist., Vol. 35, pp. 73-101, 1964.

[6] R.A. Maronna, "Robust M-estimators of multivariate location and scatter," Ann. Statist., Vol. 4, pp. 51-67, 1976.

[7] N.A. Cambell, "Mixture models and atypical values," Math. Geol., Vol. 16, pp. 465-477, 1984.

[8] Y. Jhung, “Bayesian Contextual Classification of Noise-Contaminated Multi-variate Images,” Ph.D. Dissertation, School of Electrical Engineering, Purdue University, 1994.

[9] S. Tadjudin, Classification of High Dimensional Data with Limited Training Samples, Ph.D. Thesis, School of Electrical and Computer Engineering, Purdue University, 1998 (123 pages). Available from http://dynamo.ecn.purdue.edu/~landgreb/publications.html.

[10] Swain, P. H. and S. M. Davis, eds., “Remote Sensing: The Quantitative Approach,” McGraw Hill, 1978, Chapt. 2.

[11] Hoel, P. G., Port, S. C. and Stone, C. J., “Introduction to Statistical Theory,” Houghton Mifflin Co., 1971, Chapter 3.